Appl. Comput. Harmon. Anal. 33 (2012) 44–57

Contents lists available at SciVerse ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Tree approximation with anisotropic decompositions ☆

P. Grohs

ETH Zürich, Seminar for Applied Mathematics, Rämistrasse 101, 8092 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 30 November 2010
Revised 19 June 2011
Accepted 19 September 2011
Available online 24 September 2011
Communicated by Naoki Saito

Keywords:
Shearlets
Curvelets
Tree approximation
Bit rate coding

ABSTRACT

In recent years anisotropic transforms like the shearlet or curvelet transform have received a considerable amount of interest due to their ability to efficiently capture anisotropic features in terms of nonlinear N-term approximation. In this paper we study tree-approximation properties of such transforms where the N-term approximant has to satisfy the additional constraint that the set of kept indices possesses a tree structure. The main result of this paper is that for shearlet- and related systems, this additional constraint does not deteriorate the approximation rate. As an application of our results we construct (almost) optimal encoding schemes for cartoon images.

1. Introduction

In many applications of mathematics one has to deal with piecewise smooth functions where the discontinuity arises along a smooth submanifold of the domain of definition. A particular case is given by bivariate functions which are smooth except for a smooth discontinuity curve. To give some examples of interest we mention that such functions have become widely recognized as a suitable model for image data and also arise as solutions to transport problems. It is therefore of eminent interest to come up with simple and accurate schemes to encode such data. Until a few years ago, only adaptive schemes have been available for this task, where adaptive means that one essentially has to track down the discontinuity curve and then adapt the approximation procedure to the curve [10,16,11,9] (in [3] this is called the *Lagrangian viewpoint*). In a remarkable work, in 2002 Candes and Donoho for the first time came up with a nonadaptive approximation procedure for bivariate functions which is very simple – it is defined by hard thresholding of the transform coefficients in a curvelet frame – and which possesses (almost) optimal convergence properties [4]. These results have been followed by other, similar constructions, most notably shearlets [29] and contourlets [14].

While hard thresholding of the transform coefficients gives optimal approximation rates in terms of the number of kept coefficients, there remains the question of an optimal scheme for transforming the list of kept coefficients into a sequence of bits, as would certainly be necessary for practical purposes. Questions like this can be cast in the theory of rate-distortion coding [1]. Given a model of signals to be encoded (in our case the cartoon images to be defined below) we are looking for an *encoding map* which maps a signal onto a string of bits. The maximal length of such a string over all signals is called the *runlength*. Finally, a *decoding map* is considered which maps a given bitstream to a signal. The crucial quantity which measures the performance of an encoding/decoding pair is the *distortion*, which quantifies the relation between the runlength of an encoding scheme to the maximal error in the reconstruction, see also Section 4 for a rigorous definition.

For most classes of signals there exist lower bounds on the distortion rate one can possibly achieve which allows us to use the notion of optimality. Such bounds are related to the *Kolmogorov entropy* of the signal model [31,1].

☆ This work has been funded by the European Research Council under Grant ERC Project STAHDPDE No. 247277.
E-mail address: philippgrohs@gmail.com.

1063-5203/$ – see front matter
doi:10.1016/j.acha.2011.09.004

Shearlets or curvelets provide a particularly simple encoding/decoding pair via N-term approximation: Simply keep the largest N coefficients in the frame expansion of a given signal and transform these into bits.

In attempting to assess the distortion of such an encoding scheme, the key problem that arises is that the storage cost of the indices of the kept coefficients might actually dominate the whole cost. For wavelet compression (or more general compression with orthobases), there exist several ways to remedy this problem, see for example [15,7].

We would like to focus on [7] which is based on the idea of requiring that the set of kept indices possesses a tree structure. In this way, using the fact that trees can be encoded very efficiently, optimal bit-rate codes for unit balls of Besov spaces can be constructed with wavelets.

The central problem in proving results of that kind is to show that the additional assumption on the set of retained basis (or frame) coefficients to possess a tree structure does not corrupt the best N-term approximation rate. In fact, this is the main result of [7] for the $L_p(\mathbb{R})$-approximation of Besov balls in a wavelet basis, as long as the Besov spaces lie above the 'critical embedding line'. This latter condition can be seen as a smoothness condition – it expresses a certain decay of the coefficients in the wavelet expansion of a function with scale.

The main purpose of this paper is to show analogous results for the approximation of bivariate functions with smooth discontinuity curves by anisotropic transformations based on parabolic scaling, e.g. shearlets or curvelets.

Our main result, Theorem 6 is that the additional requirement of possessing a tree structure does not deteriorate the N-term approximation rate. Since, in contrast to wavelets, curvelets and shearlets are not known to provide unconditional bases of $L_p(\mathbb{R}^2)$ for $p \neq 2$ we only consider approximation in $L_2(\mathbb{R}^2)$. In our analysis the rôle of the condition on the Besov ball to lie above the 'critical embedding line' in the results of [7] is played by Lemma 9 below which shows that a similar property holds for cartoon images and shearlet frames. Another crucial tool in our construction of efficient encoding schemes is the recent introduction of compactly supported shearlet frames [24].

We would like to mention that the scope of tree approximation is much broader than simply constructing optimal encoding schemes, since for implementational purposes it is often beneficial to store the index set as a tree.

1.1. Outline

We give an outline of this paper: Below, in Section 2 we collect various definitions and results that will be needed later on. For convenience we have decided to put a focus on the shearlet transform and therefore we explain the classical construction of a shearlet Parseval frame. We also introduce the tree structure that is inherently present in the shearlet index set. Section 3 contains our main optimality result for tree approximation. We first prove the result for the shearlet Parseval frame introduced in Section 2. Then we introduce a localization concept that allows us to transfer this result to other systems such as curvelets or different shearlet systems, compactly supported shearlets being one important example. Finally, in Section 4 we apply the results obtained in Section 3 and show how to construct a simple coding procedure which performs (almost) optimally in the sense of rate-distortion coding [1]. There, it will turn out to be crucial to use compactly supported shearlet frames as opposed to bandlimited ones. As an application we give a bound on the Kolmogorov entropy of the class $\mathcal{F}$ of cartoon images defined in Section 2.

1.2. Notation

We will use the asymptotic notation $A \lesssim B$ to indicate that A is bounded by a uniform constant times B in magnitude. If $A \lesssim B$ and $B \lesssim A$ we write $A \sim B$. For a tempered distribution f we denote by $\hat{f}$ its Fourier transform (the specific choice of normalization will not be relevant for us). The symbol $\lceil x \rceil$ denotes the smallest integer which is greater than x. We will use the symbol $|\cdot|$ in three instances: to denote the absolute value of a complex number, to denote the cardinality of a set and to denote the scale of a shearlet index (see below).

2. Preliminaries

2.1. Cartoon images

For several years it has been popular to model image data as elements of (the unit ball of) the space of functions of bounded variation or Besov spaces. For these models wavelet methods can be shown to perform optimally in the task of encoding an image [12,8]. However, this model does not fully pay tribute to the fact that an image is mostly defined by its edges, i.e. discontinuities along curves. Recently another model for so-called *cartoon images* has found a growing interest in the community. Following [4,16] we introduce the class of functions we wish to approximate. Let $\mathrm{STAR}^2(\nu)$ be the class of indicator functions χ_B of sets B with $B \subset [0,1]^2$ and ∂B a C^2-curve with curvature $\leqslant \nu$. More precisely $\mathrm{STAR}^2(\nu)$ consists of indicator functions of sets B which are (modulo translation) of the form

$$B = \left\{x \in \mathbb{R}^2 \colon |x| \leqslant \rho(\varphi),\ x = (|x|, \varphi) \text{ in polar coordinates}\right\}$$

with

$$\sup_{\varphi} \left|\rho''(\varphi)\right| \leqslant \nu, \qquad \sup_{\varphi} \left|\rho(\varphi)\right| < 1.$$

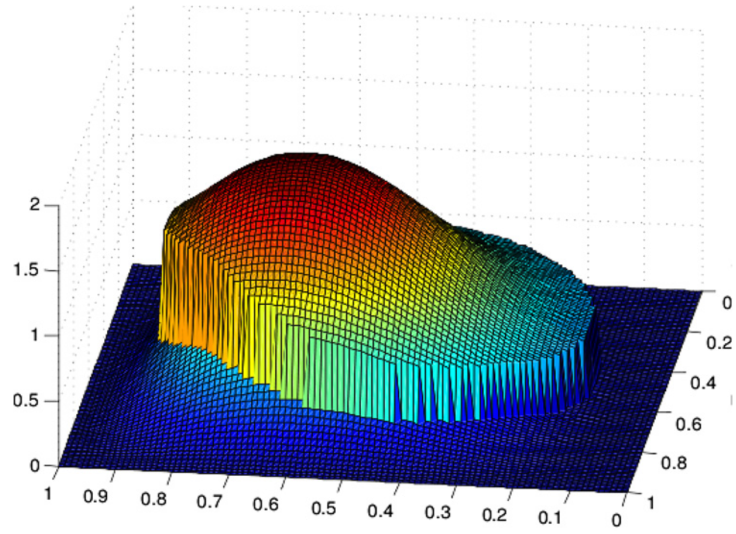

Fig. 1. A typical element of $\mathcal{F}$.

Then we define the set of *cartoon images* as

$$\mathcal{F}(\nu) := \left\{ f = f_0 + f_1 \chi_B \colon \operatorname{supp} f_i \subset [0,1]^2,\ \chi_B \in \mathrm{STAR}^2(\nu) \text{ and } \|f_0\|_{C^2}, \|f_1\|_{C^2} \leqslant 1 \right\},$$

where we write

$$\|f\|_{C^2} := \sum_{|\alpha| \leqslant 2} \left\| D^\alpha f \right\|_\infty,$$

D^α denoting the partial derivative w.r.t. $\alpha \in \mathbb{R}^2$. This definition essentially means that a function is in $\mathcal{F}(\nu)$ if it is smooth except for a C^2 discontinuity curve, see Fig. 1. Since the dependence on the parameter ν will not appear in our results, we will from now on simply write $\mathcal{F}$ instead of $\mathcal{F}(\nu)$.

This set of functions has served as a popular model for images for a while and therefore it is a crucial question how well one can approximate functions in $\mathcal{F}$. In the seminal paper [4], it has been shown that one can actually get (almost) optimal N-term approximation rates of $N^{-1}\log(N)^{3/2}$ for $\mathcal{F}$ if one expands a function in terms of a curvelet frame and keeps only the largest coefficients – the optimal achievable N-term approximation rate being N^{-1} [17]. This stands in contrast to wavelet methods which can be shown to converge only at half the rate of curvelets, namely $N^{-1/2}$. If one is willing to agree on the fact that $\mathcal{F}$ is a more realistic model for images than for instance unit balls in Besov spaces, then this shows that curvelets are superior to wavelets for the encoding of images. Despite these theoretical results, there remain several issues regarding a simple and fast implementation of a curvelet transform. Indeed, since curvelets are defined by applying rotations to various basis functions, and since it is not clear how to translate this operation to a digital grid, the actual implementations of curvelet transforms are usually not fully faithful to the continuous theory. As a remedy to this problem shearlets have been introduced in [29]. There, the operation of rotation is replaced by a shearing operation which *can* be defined on a digital grid. Moreover, the desirable approximation properties of curvelets still remain valid for shearlets. In [28] it has been shown that the approximation results for shearlets even remain valid for a more general image model which allows for piecewise C^2 discontinuity curves. Our results remain valid for this more general image model.

2.2. Shearlets

The main goal of this paper is to show that $\mathcal{F}$ can still be (almost) optimally approximated if one imposes the additional constraint on the kept indices to form a tree. This is highly desirable for deriving efficient coding procedures as well as certain implementational issues. In recent years several systems which are well-adapted to the image class $\mathcal{F}$ have been developed, among them we mention curvelets [4], shearlets [29] and contourlets [14]. For our purposes shearlets stand out for the following reasons:

- Shearlets are defined over a uniform grid which makes it much easier to define a suitable parent–child relation on the index set,
- there exist constructions of compactly supported shearlet frames [24], a property that will turn out essential for constructing (almost) optimal coding schemes.

The development of shearlets has by now reached a somewhat mature state, the reader interested in the current state of the art is referred to the survey papers [25,26].

We now describe the main definitions and notation related to shearlets. First we need the concept of a *frame* [6].

Definition 1. A system $\Psi = (\psi_\omega)_{\omega\in\Omega}$ of elements ψ_ω in a Hilbert space $\mathcal{H}$, indexed by a countable index set Ω is called a *frame* if

$$\sum_{\omega\in\Omega} \left|\langle f, \psi_\omega\rangle\right|^2 \sim \|f\|_{\mathcal{H}}^2 \quad \text{for all } f \in \mathcal{H}. \tag{1}$$

Equivalent to (1) is the following condition:

$$\inf_{\sum_{\omega\in\Omega} c_\omega\psi_\omega = f} \sum_{\omega\in\Omega} |c_\omega|^2 \sim \|f\|_{\mathcal{H}}^2. \tag{2}$$

For a general frame Ψ there exists a dual frame $\tilde{\Psi} = (\tilde{\psi}_\omega)_{\omega\in\Omega}$, indexed over the same index set, such that the representation

$$f = \sum_{\omega\in\Omega} \langle f, \psi_\omega\rangle \tilde{\psi}_\omega \tag{3}$$

holds in $\mathcal{H}$. If (1) holds with '=' instead of '$\sim$', then Ψ is called a *Parseval frame*. In this case we have the representation

$$f = \sum_{\omega\in\Omega} \langle f, \psi_\omega\rangle \psi_\omega \tag{4}$$

in $\mathcal{H}$.

In the following we will only be interested in frames of the Hilbert space $L_2(\mathbb{R}^2)$. There exist two principally different constructions of shearlet frames for $L_2(\mathbb{R}^2)$: bandlimited frames, introduced in [29] and compactly supported frames, introduced in [24]. We briefly describe both of them, starting with the bandlimited construction. Shearlets are generally built from a finite set of basis functions using the operations of translation, anisotropic dilation, described by the matrices

$$A_0 := \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad A_1 := \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix},$$

and shearing, described by the matrices

$$B_0 := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B_1 := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

2.2.1. Bandlimited shearlets

We now follow [21], where specific bandlimited functions $\varphi, \psi^{(0)}, \psi^{(1)}$ are constructed in the Fourier domain from which a Parseval frame is built. The construction of $\psi^{(0)}$ uses specific functions $\hat{V}, \hat{W}$ with $\operatorname{supp}\hat{V} \subset [-1, 1]$ and $\operatorname{supp}\hat{W} \subset [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$ satisfying certain discrete Calderòn-type conditions, see [21, Eqs. (1.6), (1.7)] and also [23, Section 5.2.1]. Now the function $\psi^{(0)}$ is defined via its Fourier transform as

$$\hat{\psi}^{(0)}(\xi) := \hat{W}(\xi_1)\hat{V}\left(\frac{\xi_2}{\xi_1}\right). \tag{5}$$

The construction of $\psi^{(1)}$ is similar with the rôles of the coordinate axes reversed. Finally, an appropriate φ with $\operatorname{supp}\hat{\varphi} \subset [-\frac{1}{8}, \frac{1}{8}]^2$ is constructed. Then, in [19, Theorem 2.1] it is shown that with

$$\sigma_{(j,l,k,d)} := 2^{3j/2}\psi^{(d)}\big(B_d^l A_d^j \cdot - k\big), \qquad \sigma_k := \varphi(\cdot - k),$$

the system

$$\Sigma := \{\sigma_k \colon k \in \mathbb{Z}\} \cup \big\{\sigma_{(j,k,l,d)} \colon j \geqslant 0,\ -2^j \leqslant l \leqslant 2^j - 1,\ k \in \mathbb{Z}^2,\ d = 0, 1\big\}$$

constitutes a Parseval frame for $L_2(\mathbb{R}^2)$.

Remark 2. Actually this statement is not quite true: one has to slightly modify the elements corresponding to indices with $l = -2^j$ and $l = 2^j - 1$ in order to obtain a Parseval frame, compare [19]. We will not go into this issue.

2.2.2. Compactly supported shearlets

The Parseval frame construction has the nice property that the reconstruction of a function $f \in L_2(\mathbb{R}^2)$ is particularly simple using the expansion (4). On the other hand, it possesses the drawback of being composed of noncompactly supported functions. For several applications it is desirable to have a frame at hand that is constructed of compactly supported functions – as a matter of fact we will crucially require this property in our construction of optimal codes below. To satisfy this need, compactly supported shearlet frames have been presented in [24]. We briefly describe the construction which again is based on three functions $\varphi', \psi'^{(0)}, \psi'^{(1)}$ – this time compactly supported. Without going into the (considerable) details we mention that for all $R \in \mathbb{R}_+$ there exist constructions of compactly supported functions $\varphi', \psi'^{(0)}, \psi'^{(1)} \in C^R(\mathbb{R}^2)$ and a sampling constant $\delta > 0$ such that with

$$\sigma'_{(j,l,k,d)} := 2^{3j/2}\psi'^{(d)}\big(B_d^l A_d^j \cdot - \delta k\big), \qquad \sigma'_k := \varphi'(\cdot - \delta k),$$

the system

$$\Sigma' := \big\{\sigma'_k \colon k \in \mathbb{Z}\big\} \cup \big\{\sigma'_{(j,k,l,d)} \colon j \geqslant 0,\ -2^j \leqslant l \leqslant 2^j - 1,\ k \in \mathbb{Z}^2,\ d = 0, 1\big\} \tag{6}$$

constitutes a frame for $L_2(\mathbb{R}^2)$. The function $\psi'^{(0)}$ can be constructed as a tensor product function

$$\psi'^{(0)}(x_1, x_2) = W(x_1)V(x_2)$$

of a wavelet W in the first coordinate and a scaling function V in the second coordinate, both associated with a carefully calibrated family of filters, see [24, Theorem 4.7] (compare this separable construction with the bandlimited construction (5)). Furthermore, the function $\psi'^{(0)}$ can be constructed to satisfy the moment condition

$$\big|\hat{\psi}'^{(0)}(\xi)\big| \lesssim |\xi_1|^R, \tag{7}$$

see [24, Proposition 4.6]. The construction of $\psi'^{(1)}$ goes along the same lines with the two coordinates reversed, and the low-pass function φ' can be taken as $\varphi'(x_1, x_2) = V(x_1)V(x_2)$. In the remainder we shall denote a frame of compactly supported shearlets with Σ', keeping in mind that it is possible to construct the basis functions to be arbitrarily smooth with (7) holding true for R arbitrarily large.

2.2.3. Tree structure

With

$$\Lambda_{-1} := \mathbb{Z}^2 \quad \text{and} \quad \Lambda_j := \big\{(j, l, k, d) \colon -2^j \leqslant l \leqslant 2^j - 1,\ k \in \mathbb{Z}^2,\ d = 0, 1\big\},$$

we define the shearlet index set $\Lambda = \dot{\bigcup}_{j \geqslant -1} \Lambda_j$ and get the representation (valid in $L_2(\mathbb{R}^2)$)

$$f = \sum_{\lambda \in \Lambda} \langle f, \sigma_\lambda\rangle \sigma_\lambda \tag{8}$$

in the bandlimited case, and

$$f = \sum_{\lambda \in \Lambda} \langle f, \sigma'_\lambda\rangle \tilde{\sigma}'_\lambda \tag{9}$$

in the compactly supported case – $\tilde{\Sigma}' = (\tilde{\sigma}'_\lambda)_{\lambda \in \Lambda}$ denoting a dual frame of Σ'. The shearlet index set Λ carries a natural tree structure which we will now describe. For an index $\lambda \in \Lambda$ we write $|\lambda|$ to denote the unique integer j with $\lambda \in \Lambda_j$. Further we write

$$\mathcal{E}_0 := \big\{(0,0), (1,0), (2,0), (3,0), (0,1), (1,1), (2,1), (3,1)\big\}$$

and

$$\mathcal{E}_1 := \big\{(0,0), (0,1), (0,2), (0,3), (1,0), (1,1), (1,2), (1,3)\big\}.$$

Definition 3. An index $(0, l, k, d) \in \Lambda_0$ is called a *child* of $m \in \Lambda_{-1}$ if $k = B_d^l m$. An index $(j, l, k, d) \in \Lambda_j$ is called a *child* of (j', l', k', d') if $d = d'$, $j = j' + 1$, $l \in \{2l', 2l' + 1\}$ and $k \in B_d^{\lceil l/2\rceil - l'}(A_d k' + \mathcal{E}_d)$ (see Fig. 2). We can transitively extend this relation and write $\lambda' \preccurlyeq \lambda$ if either $\lambda = \lambda'$ or λ' is a child of λ.

Now, having defined a natural parent–child relation on the set of shearlet indices, we are ready to state precisely what we understand as a tree.

Definition 4. Every $\lambda \in \Lambda_j$ possesses a *unique* parent in Λ_{j-1}, $j \geqslant 0$ and 16 children in Λ_{j+1} for $j \geqslant 0$ and 4 children for $j = -1$. We call a subset $\mathcal{T} \subset \Lambda$ a *tree* if for every $\lambda \in \mathcal{T}$ also its parent is contained in $\mathcal{T}$.

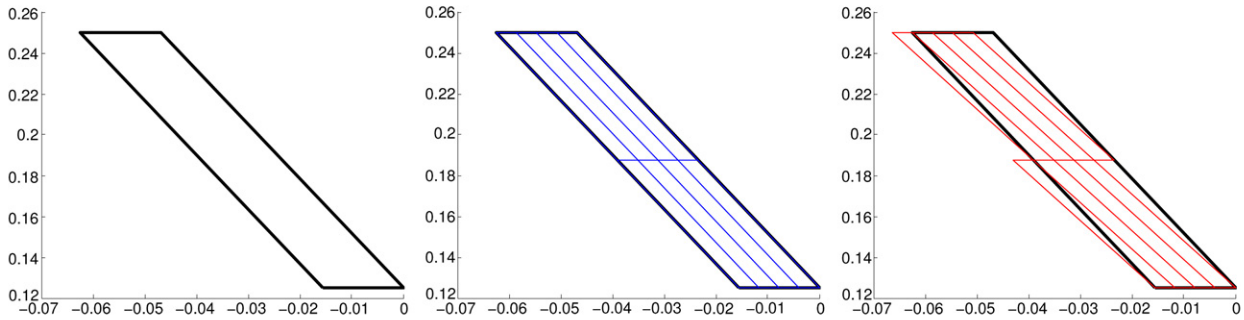

Fig. 2. Left: Essential support of σ_λ with $j=3$, $l=3$, $d=0$, $k=(2,1)$. Middle: Essential support of its children with $l=6$. Right: Essential support of its children with $l=7$.

3. The optimality result

This section contains our main result, namely that the best N-term approximation rate can be retained by requiring that the set of kept indices forms a tree. Let us first give some motivation. We will focus on the bandlimited shearlet tight frame Σ introduced above in Section 2.2.1. Having this construction at hand it is a central question how the shearlet frame is adapted to the structure of the image model $\mathcal{F}$. Mathematically, such questions can be formalized by the theory of nonlinear approximation [12].

From the viewpoint of nonlinear approximation the central problem is how well an arbitrary $f \in \mathcal{F}$ can be approximated in $L_2(\mathbb{R}^2)$ by keeping only $N \in \mathbb{N}$ coefficients in the expansion (8).

More formally one can introduce the nonlinear approximation spaces

$$\Sigma_N := \left\{ \sum_{\lambda\in\mathcal{I}} c_\lambda \sigma_\lambda \colon\ |\mathcal{I}| \leqslant N \right\}$$

and study the asymptotic behavior of the *best N-term approximation error*

$$s_N(f) := \inf_{g\in\Sigma_N} \|f-g\|_2$$

for $f \in \mathcal{F}$. Often there is a precise limit as to what one can expect from the behavior of this approximation error. Indeed it can be shown (by giving lower bounds on the metric entropy of $\mathcal{F}$ [17]) that no frame can deliver a better rate than

$$\sup_{f\in\mathcal{F}} s_N(f) \lesssim N^{-1},$$

at least under very weak additional assumptions on how a best N-term approximation can be found. It is not known whether this optimal rate can be achieved. Nevertheless, there exists the following remarkable result:

Theorem 5. *(See [21].) We have for all $\varepsilon > 0$ the approximation*

$$\sup_{f\in\mathcal{F}} s_N(f) := \inf_{g\in\Sigma_N} \|f-g\|_2 \lesssim N^{-1+\varepsilon}.$$

In fact, more is true: the best N-term approximation of a given $f \in \mathcal{F}$ can simply be computed by keeping the N largest coefficients in the frame expansion (8). Theorem 5 has been inspired by the analogous result for curvelets in [4]. In [27] this result has been extended to compactly supported shearlets. Our main result is that we can still get close-to-optimal N-term approximation performance if we only keep index sets forming a tree. The proofs utilize various concepts from nonlinear approximation [12] and wavelet tree approximation [7].

We define the approximation spaces

$$\Sigma_N^t := \left\{ \sum_{\lambda\in\mathcal{T}} c_\lambda \sigma_\lambda \colon\ \mathcal{T} \text{ is tree, and } |\mathcal{T}| \leqslant N \right\}$$

and would like to answer the following question: *What is the asymptotic rate of the error*

$$t_N(f) := \inf_{g\in\Sigma_N^t} \|f-g\|_2,$$

where f ranges in $\mathcal{F}$?

The goal is to show the analogous statement to Theorem 5 for tree approximation with shearlets, namely:

Theorem 6. *We have for all $\varepsilon > 0$ the approximation*

$$\sup_{f\in\mathcal{F}} t_N(f) := \sup_{f\in\mathcal{F}} \inf_{g\in\Sigma_N^t} \|f-g\|_2 \lesssim N^{-1+\varepsilon}.$$

The basic strategy to prove this result is to count the number of frame coefficients exceeding a threshold $\eta > 0$. To this end we let

$$\Lambda(f,\eta) := \{\lambda \in \Lambda\colon\ |\langle f, \sigma_\lambda\rangle| \geqslant \eta\}, \qquad \Lambda_j(f,\eta) := \Lambda(f,\eta) \cap \Lambda_j.$$

Once precise estimates for the asymptotics of $\Lambda(f,\eta)$ are known for $\eta \to 0$, it is possible to quantify the N-term approximation rate. This can be made precise using the notion of weak l_p spaces (a.k.a. Lorentz spaces) which are defined on countable complex-valued sequences $\Gamma = (\gamma_n)_{n\in\mathbb{Z}}$ via the quasinorm

$$\|\Gamma\|_{wl_p} := \sup_{n>0} n^{1/p} |\gamma_n^*|, \tag{10}$$

where $\Gamma^* = (\gamma_n^*)_{n\in\mathbb{N}}$ denotes *decreasing rearrangement* of Γ [13]. To see how this is related to best N-term approximation and counting indices in $\Lambda(f,\eta)$ we remark that

$$\|\Gamma\|_{wl_p} \sim \sup_{\eta>0} \eta^p \left|\{n\colon\ |\gamma_n| \geqslant \eta\}\right|,$$

as can easily be shown [13].

We want to approximate only with N-term approximations where the set of kept indices forms a tree. To this end we define $\mathcal{T}(f,\eta)$ to be the smallest tree containing $\Lambda(f,\eta)$ and $\mathcal{T}_j(f,\eta) := \mathcal{T}(f,\eta) \cap \Lambda_j$. Clearly, we have $\mathcal{T}(f,\eta) \subset \mathcal{T}(f,\eta')$ for $\eta \geqslant \eta'$.

Before we can prove our main result we first need to state a series of auxiliary results.

Our starting point is a result of Guo and Labate in [21] where the unit square is partitioned into dyadic squares of sidelength $\sim 2^{-j}$ and f is localized onto each such square using a smooth partition of unity. We denote by $\mathcal{Q}$ the collection of all these squares tiling the unit square. Further, we denote the localization of f onto a dyadic square Q by f_Q and consider the coefficient sequence

$$\Gamma_Q := (\langle f_Q, \sigma_\lambda\rangle)_{\lambda\in\Lambda_j}.$$

There are two different types of elements in $\mathcal{Q}$: Those which intersect the singularity curve and those which do not. We call the collection of squares of the first type $\mathcal{Q}^0$ and the collection of squares of the latter type $\mathcal{Q}^1$. We shall now use a key result that has been proven in [21].

Lemma 7. *(See [21, Theorems 1.3, 1.4].) For $Q \in \mathcal{Q}^0$ we have*

$$\|\Gamma_Q\|_{wl_{2/3}} \lesssim 2^{-3j/2}.$$

For $Q \in \mathcal{Q}^1$ we have

$$\|\Gamma_Q\|_{2/3} \lesssim 2^{-3j}.$$

The implicit constants are independent of scale j.

Using this result we can prove

Corollary 8. *For all $\varepsilon > 0$ and $Q \in \mathcal{Q}^0$ we have*

$$\|\Gamma_Q\|_{2/3+\varepsilon} \lesssim 2^{-3j/2}.$$

For $Q \in \mathcal{Q}^1$ we have

$$\|\Gamma_Q\|_{2/3+\varepsilon} \lesssim 2^{-3j}.$$

The implicit constants are independent of scale j.

Proof. Let $Q \in \mathcal{Q}_0$ and denote by $\Gamma_Q^* = ((\gamma_Q^*)_n)_{n\in\mathbb{N}}$ the decreasing rearrangement of Γ_Q. By Lemma 7 and the definition (10) of the weak l_p quasinorm we have

$$|(\gamma_Q^*)_n| \leqslant \|\Gamma_Q\|_{wl_{2/3}} n^{-3/2} \lesssim 2^{-3j/2} n^{-3/2}.$$

Therefore

$$\|\Gamma_Q\|_{2/3+\varepsilon}^{2/3+\varepsilon} = \sum_{n\in\mathbb{N}} \left|(\gamma_Q^*)_n\right|^{2/3+\varepsilon} \lesssim 2^{-(3j/2)(2/3+\varepsilon)} \sum_{n\in\mathbb{N}} n^{-1-3\varepsilon/2} \lesssim 2^{-(3j/2)(2/3+\varepsilon)}.$$

The case $Q \in \mathcal{Q}_1$ is the same. $\square$

Lemma 9. *Let $f \in \mathcal{F}$. Then for any $\varepsilon > 0$ there exists $\delta > 0$ such that we have the estimate*

$$\left\|\left(\langle f, \sigma_\lambda\rangle\right)_{\lambda\in\Lambda_j}\right\|_{2/3+\varepsilon} \lesssim 2^{-\delta j} \tag{11}$$

and hence

$$\left|\Lambda_j(f,\eta)\right| \lesssim 2^{-\delta j}\eta^{-(2/3+\varepsilon)} \quad \textit{for all } j \geqslant 0. \tag{12}$$

Proof. We use the p-triangle inequality for $p = 2/3+\varepsilon$ with Corollary 8 and compute

$$\left\|\left(\langle f, \sigma_\lambda\rangle\right)_{\Lambda_j}\right\|_{2/3+\varepsilon}^{2/3+\varepsilon} \leqslant \sum_{Q\in\mathcal{Q}^0} \|\Gamma_Q\|_{2/3+\varepsilon}^{2/3+\varepsilon} + \sum_{Q\in\mathcal{Q}^1} \|\Gamma_Q\|_{2/3+\varepsilon}^{2/3+\varepsilon} \lesssim 2^j 2^{-j-3j\varepsilon/2} + 4^j 2^{-2j-3\varepsilon} \lesssim 2^{-3j\varepsilon/2}.$$

We have used the fact that $|\mathcal{Q}^0| \lesssim 2^j$ and $|\mathcal{Q}^1| \lesssim 4^j$. It is a well-known fact that for general sequences (γ_n) we have the inequality

$$\sup_{\eta>0}\left|\left\{n:\ |\gamma_n| \geqslant \eta\right\}\right|\eta^p \leqslant \left\|(\gamma_n)\right\|_p.$$

With $p = 2/3+\varepsilon$ this implies that

$$\sup_{\eta>0}\left|\left\{\lambda \in \Lambda_j:\ \left|\langle f, \sigma_\lambda\rangle\right| \geqslant \eta\right\}\right|\eta^p \lesssim 2^{-\delta j},$$

where $\delta := \frac{3\varepsilon/2}{2/3+\varepsilon} > 0$. This proves the desired statement. $\square$

Lemma 9 allows us to count the elements of the set $\mathcal{T}(f,\eta)$ as η goes to zero.

Lemma 10. *We have for any $\varepsilon > 0$*

$$\left|\mathcal{T}(f,\eta)\right| \lesssim \eta^{-(2/3+\varepsilon)}. \tag{13}$$

Proof. We show the estimate

$$\left|\mathcal{T}_j(f,\eta)\right| \lesssim 2^{-\delta j}\eta^{-(2/3+\varepsilon)},$$

which implies the desired result. Every element in $\mathcal{T}_j(f,\eta)$ is either in $\Lambda_j(f,\eta)$ or it is the unique parent of some $\lambda' \in \Lambda_{j'}(f,\eta)$, $j' > j$. Therefore we have

$$\left|\mathcal{T}_j(f,\eta)\right| \leqslant \sum_{j'\geqslant j}\left|\Lambda_{j'}(f,\eta)\right|. \tag{14}$$

From (9) we know that for $f \in \mathcal{F}$ with some $\delta > 0$ depending only on ε we have the estimate

$$\left|\Lambda_{j'}(f,\eta)\right| \lesssim 2^{-\delta j'}\eta^{-(2/3+\varepsilon)}.$$

This implies the desired result. $\square$

Having a bound for the cardinality of $\mathcal{T}(f,\eta)$ we now approximate f by only keeping the indices in $\mathcal{T}(f,\eta)$.

Definition 11. Define the tree approximant

$$\mathcal{S}(f,\eta) := \sum_{\lambda\in\mathcal{T}(f,\eta)} \langle f, \sigma_\lambda\rangle\sigma_\lambda.$$

Lemma 12. *For any $\varepsilon > 0$ we have the approximation rate*

$$\left\| f - \mathcal{S}(f, \eta) \right\|_2 \lesssim \eta^{2/3-\varepsilon}, \tag{15}$$

uniformly over $f \in \mathcal{F}$.

Proof. Define

$$u_l := \sum_{\lambda \in \mathcal{T}(f, 2^{-(l+1)}\eta) \setminus \mathcal{T}(f, 2^{-l}\eta)} \langle f, \sigma_\lambda \rangle \sigma_\lambda.$$

Due to the frame property of Σ (in particular because of (2)) we can estimate for any $\varepsilon' > 0$

$$\|u_l\|_2 \leqslant \left(\sum_{\lambda \in \mathcal{T}(f, 2^{-(l+1)}\eta) \setminus \mathcal{T}(f, 2^{-l}\eta)} \left| \langle f, \sigma_\lambda \rangle \right|^2 \right)^{1/2} \leqslant 2^{-l}\eta \left(\left| \mathcal{T}\left(f, 2^{-l-1}\eta\right) \right| \right)^{1/2} \lesssim 2^{-l(2/3-\varepsilon'/2)} \eta^{2/3-\varepsilon'/2}.$$

We have used (13) in the last estimate. Now, since

$$\left\| f - \mathcal{S}(f, \eta) \right\|_2 \leqslant \sum_l \|u_l\| \lesssim \eta^{2/3-\varepsilon'/2}$$

we arrive at the desired result by setting $\varepsilon' := 2\varepsilon$. $\square$

We are finally ready to formulate and prove of our main result.

Proof of Theorem 6. Let $\eta^{2/3+\varepsilon/2} := N^{-1}$ with $\varepsilon > 0$ fixed. Then by (13) we have

$$\left| \mathcal{T}(f, \eta) \right| \lesssim \eta^{-(2/3+\varepsilon/2)} = N$$

and by (15) we have

$$\left\| f - \mathcal{S}(f, \eta) \right\|_2 \lesssim \eta^{2/3-\varepsilon/6} = N^{-\frac{2/3-\varepsilon/6}{2/3+\varepsilon/2}} = N^{-1} N^{\frac{\varepsilon/2+\varepsilon/6}{2/3+\varepsilon/2}} \leqslant N^{-1} N^{\varepsilon \frac{2/3}{2/3}} = N^{-1+\varepsilon}. \qquad \square$$

Remark 13. Actually, one can show a slightly better result than Theorem 6 by noting that due to the uniform boundedness of functions in $\mathcal{F}$ and the structure of the shearlet frame we can estimate

$$\left| \langle f, \sigma_\lambda \rangle \right| \leqslant \|f\|_\infty \|\sigma_\lambda\|_1 \lesssim 2^{-3|\lambda|/2}, \tag{16}$$

which means that in (14) we only have to sum up to $j, j' \lesssim \log(\eta^{-1})$ (otherwise, by (16), the coefficients would be smaller than η). This yields the better estimate

$$\left| \mathcal{T}(f, \eta) \right| \lesssim \eta^{-2/3} \log\left(\eta^{-1}\right)^2. \tag{17}$$

Using (17) one can go on to show that

$$\sup_{f \in \mathcal{F}} t_N(f) \lesssim N^{-1} \log(N)^3.$$

We do not know if this can be improved.

Remark 14. In [28] it is shown that shearlets retain their best N-term approximation rate if the singularity curve of f is allowed to possess a finite number of kinks. Our results also remain valid for this more general image model. This can be seen by examining the proofs of [28] which also goes by counting the quantities $\Lambda(f, \eta)$.

3.1. Optimal tree approximation for other systems

In the proof of our main theorem we have assumed that we are given a tight frame of bandlimited shearlets in order to make use of the results in [22]. Naturally, the question arises whether these assumptions are crucial. Actually, they are not. We can get the same approximation rate for tree approximation with any shearlet or curvelet system provided that the underlying basis functions are sufficiently smooth, sufficiently localized in space and possess sufficiently many anisotropic vanishing moments. The main idea is to build on Theorem 6 and to examine the cross Gramian matrix $(\langle \sigma_\lambda, \theta_\mu \rangle)$ between the bandlimited shearlet frame Σ and another frame $\Theta = (\theta_\mu)_{\mu \in M}$, which might be another shearlet or even a curvelet frame. Before we can state our main result we need some definitions.

Consider two hierarchical index sets M, M', meaning that we have a disjoint union $M = \dot{\bigcup}_{j \geqslant 0} M_j$ and $M' = \dot{\bigcup}_{j \geqslant 0} M'_j$. We can associate to each $\mu \in M$ its scale $|\mu|$ which is the unique index j such that $\mu \in M_j$. The same can be done with M'. We now define a localization concept that will turn out to be useful for us.

Definition 15. Two systems $\Theta := (\theta_\mu)_{\mu\in M}$, $\Theta' := (\theta'_{\mu'})_{\mu'\in M'}$ are called *L-localized*, $L > 0$, if their cross-Gramian is almost diagonal in the sense that

$$\big|\langle\theta_\mu, \theta'_{\mu'}\rangle\big| \lesssim 4^{-L||\mu|-|\mu'||}\omega\big(\mu, \mu'\big)^{-L},$$

where $\omega : M \times M' \to \mathbb{R}_+$ is a distance function satisfying

$$\sup_{\mu\in M_j} \sum_{\mu'\in M'_{j'}} \omega\big(\mu, \mu'\big)^{-2} \lesssim 4^{2|j-j'|} \tag{18}$$

with the implicit constant independent of j, j'.

Lemma 16. *Assume that $\Theta = (\theta_\mu)_{\mu\in M}$ is L-localized with the bandlimited shearlet frame Σ and $L > 3$. Then for any $f \in \mathcal{F}$ and for any $\varepsilon > 0$ there exists $\delta > 0$ (depending only on ε) such that*

$$\big\|\big(\langle f, \theta_\mu\rangle\big)_{\mu\in M_j}\big\|_{2/3+\varepsilon} \lesssim 2^{-\delta j}. \tag{19}$$

Proof. Let $f \in \mathcal{F}$ and $\varepsilon > 0$. Then by Lemma 9 we have with $c_\lambda := \langle f, 2^{\delta|\lambda|}\sigma_\lambda\rangle$ and $\delta > 0$ small that

$$f = \sum_{\lambda\in\Lambda} c_\lambda 2^{-\delta|\lambda|}\sigma_\lambda, \tag{20}$$

where $\|(c_\lambda)_{\lambda\in\Lambda}\|_p < \infty$ and $p = 2/3 + \varepsilon$. We want to show that

$$\big\|\big(2^{\delta|\mu|}\langle f, \theta_\mu\rangle\big)_{\mu\in M}\big\|_p < \infty \tag{21}$$

which implies the desired claim: indeed, (21) implies that

$$\sum_{j\geqslant 0} 2^{p\delta j}\big\|\big(\langle f, \theta_\mu\rangle\big)_{\mu\in M_j}\big\|_p^p < \infty$$

which implies that

$$2^{p\delta j}\big\|\big(\langle f, \theta_\mu\rangle\big)_{\mu\in M_j}\big\|_p^p \lesssim 1,$$

which is what we want.

Clearly (21) follows if we can establish that the mapping $(c_\lambda)_{\lambda\in\Lambda} \mapsto (2^{\delta|\mu|}\langle f, \theta_\mu\rangle)_{\mu\in M}$ is bounded in l_p. The matrix of this mapping is given by

$$\big(\big\langle 2^{\delta|\mu|}\theta_\mu, 2^{-\delta|\lambda|}\sigma_\lambda\big\rangle\big)_{\lambda\in\Lambda,\mu\in M}$$

and therefore in view of Schur's lemma we need to show that

$$\sup_{\lambda\in\Lambda} \sum_{\mu\in M} \big|\big\langle 2^{-\delta|\mu|}\theta_\mu, 2^{\delta|\lambda|}\sigma_\lambda\big\rangle\big|^p < \infty. \tag{22}$$

Using the localization property of Θ we estimate

$$\begin{aligned}
\sup_{\lambda\in\Lambda} \sum_{\mu\in M} \big|\big\langle 2^{-\delta|\mu|}\theta_\mu, 2^{\delta|\lambda|}\sigma_\lambda\big\rangle\big|^p &= \sup_{\lambda\in\Lambda} \sum_{j\geqslant 0} \sum_{\mu\in M_j} \big|\big\langle 2^{-\delta|\mu|}\theta_\mu, 2^{\delta|\lambda|}\sigma_\lambda\big\rangle\big|^p \\
&\leqslant \sup_{\lambda\in\Lambda} \sum_{j\geqslant 0} 2^{p\delta||\lambda|-j|} \sum_{\mu\in M_j} \big|\langle\theta_\mu, \sigma_\lambda\rangle\big|^p \\
&\leqslant \sup_{\lambda\in\Lambda} \sum_{j\geqslant 0} 2^{p\delta||\lambda|-j|} \sum_{\mu\in M_j} 4^{-Lp||\lambda|-j|}\omega(\lambda, \mu)^{-Lp} \\
&\leqslant \sup_{\lambda\in\Lambda} \sum_{j\geqslant 0} 2^{p\delta||\lambda|-j|}4^{(2-Lp)||\lambda|-j|} \\
&= \sup_{\lambda\in\Lambda} \sum_{j\geqslant 0} 4^{(2+\delta p/2-pL)||\lambda|-j|} < \infty,
\end{aligned}$$

whenever $L > 3$ and δ sufficiently small. $\square$

Remark 17. It is certainly possible to require instead of (18) that

$$\sup_{\mu\in M_j}\sum_{\mu'\in M'_{j'}}\omega(\mu,\mu')^{-\alpha}\lesssim 4^{\beta|j-j'|}\tag{23}$$

for some $\alpha,\beta>0$. The conclusion of Lemma 16 would still hold, possibly with another constant than 3. The reason why we chose $\alpha=\beta=2$ is simply that for this choice, for many anisotropic frame decompositions, condition (18) can be verified.

We now assume that the system $\Theta=(\theta_\mu)_{\mu\in M}$ constitutes a frame for $L_2(\mathbb{R}^2)$, i.e.

$$\|f\|_2^2\sim\sum_{\mu\in M}\left|\langle f,\theta_\mu\rangle\right|^2.$$

Consider a dual frame $\tilde{\Theta}=(\tilde{\theta}_\mu)_{\mu\in M}$ with

$$f=\sum_{\mu\in M}\langle f,\theta_\mu\rangle\tilde{\theta}_\mu.$$

Having a tree structure on M we can define the set

$$\bar{\Sigma}_N^t:=\left\{\sum_{\mu\in\mathcal{T}}c_\mu\tilde{\theta}_\mu:\ \mathcal{T}\subset M\text{ is tree, and }|\mathcal{T}|\leqslant N\right\}$$

and consider the quantity

$$\bar{t}_N(f):=\inf_{g\in\bar{\Sigma}_N^t}\|f-g\|_2.$$

Theorem 18. *Assume that Θ constitutes a frame for $L_2(\mathbb{R}^2)$ and that the index set M possesses a tree structure. Assume moreover that Σ,Θ are L-localized with $L>3$. Then the conclusion of Theorem 6 holds with Σ replaced by Θ, meaning that*

$$\sup_{f\in\mathcal{F}}\bar{t}_N(f)\lesssim N^{-1+\varepsilon}\tag{24}$$

for all $\varepsilon>0$.

Proof. The proof goes by repeating the arguments leading to Theorem 6 and using the frame property of Θ and (19). Observe that the structure of the dual frame $\tilde{\Theta}$ is irrelevant for our argument to work, important is that a tree structure can be defined on M and the fact that Θ is a frame so that (2) holds true. $\square$

The reason why Theorem 18 is interesting, is that a number of anisotropic systems are localized with Σ and therefore possess the same approximation rates.

Example 19. We give some examples of systems Θ which are L-localized with Σ (without proof): Arbitrary systems of curvelet molecules of sufficient regularity are L-localized with Σ and $L>3$, see [2] for the definition and [18] for other results in this direction. Another example is given by the tight frame Φ_J constructed in [5, Section 5.2].

In this paper we would like to focus on systems Θ of *shearlet molecules* as defined in [22]:

Definition 20. A system $\Theta=((m_\lambda)_{\lambda\in\Lambda})$ of functions is called a *system of shearlet molecules of regularity R* if we can write

$$m_\lambda(\cdot)=2^{3j/2}a^{(\lambda)}\left(B_d^jA_d^l\cdot-\delta k\right),\quad \lambda=(j,l,k,d)\in\Lambda$$

with a *sampling constant* $\delta>0\in\mathbb{R}$ and functions $a^{(\lambda)}$ satisfying

$$\left|D^\mu a^{(\lambda)}(\cdot)\right|\lesssim\left(1+|\cdot|\right)^{-P}\quad\text{for all }\mu\in\mathbb{N}^2,\ |\mu|\leqslant R,\ P\in\mathbb{N}\tag{25}$$

and

$$\left|\hat{a}^{(\lambda)}(\xi)\right|\lesssim\left(4^{-j}+|\xi_{1+d}|\right)^R\left(1+|\xi|\right)^{-R},\quad \xi=(\xi_1,\xi_2)\in\mathbb{R}^2.\tag{26}$$

For a system Θ with sampling constant δ we write

$$x_\lambda:=A_d^{-j}B_d^{-l}\delta k.\tag{27}$$

We also write e_λ for the unit vector $(\cos(\theta_\lambda), \sin(\theta_\lambda))$, where $\theta_\lambda := \arctan(2^{-j}l)$. By examining the construction in [21] it is easy to see that the Parseval frame Σ is a system of shearlet molecules of arbitrary regularity. We now define a notion of distance between two indices of two (possibly different) systems of shearlet molecules. This definition follows [22] which in turn is based on [4,32].

Definition 21. We define a distance $\omega : \Lambda \times \Lambda \to \mathbb{R}_+$ between shearlet indices via

$$\omega(\lambda, \lambda') := \left(1 + 4^{\min(\lambda,\lambda')} d(\lambda, \lambda')\right),$$

where

$$d(\lambda, \lambda') := \left|2^j l - 2^{-j'} l'\right|^2 + |x_\lambda - x_{\lambda'}|^2 + \left|\langle e_\lambda, x_\lambda - x_{\lambda'}\rangle\right|.$$

It is not difficult to see that this distance satisfies (18), as shown in [22] (see also [4]). With respect to this distance, any two systems of sufficient regularity are almost orthogonal as shown in [22, Theorems 1.3 and 1.4]:

Theorem 22. *For any $L > 0$, there exists $R > 0$ such that any two systems Θ, Θ' of shearlet molecules with regularity R satisfy*

$$\left|\langle m_\lambda, m'_{\lambda'}\rangle\right| \lesssim 4^{-L||\lambda|-|\lambda'||} \omega(\lambda, \lambda')^{-L}. \tag{28}$$

In particular, Θ, Θ' are L-localized.

In particular this implies by Theorem 18 that any system of shearlet molecules of sufficient regularity forming a frame satisfies the same tree approximation rate as the bandlimited frame Σ.

Corollary 23. *There exists $R_0 > 0$ such that for all systems Θ of shearlet molecules of regularity $R > R_0$, which also form a frame for $L_2(\mathbb{R}^2)$, the conclusion of Theorem* 6 *is valid.*

Of course it would be possible to make the dependence of R on L in Theorem 22 explicit and to compute R_0, but that would be beyond the scope of this paper. Rather we would like to single out a particular system of shearlet molecules that will turn out useful in constructing encoding schemes, namely the construction given in [24] of compactly supported shearlet frames Σ' which we described in Section 2.2.2. By appropriate choice of the generators $\psi'^{(0)}, \psi'^{(1)}, \varphi'$, the system Σ' is a system of shearlet molecules of regularity $> R_0$ (compare (7) and the discussion in Section 2.2.2), and therefore, by Theorem 18, the conclusion of Theorem 6 is valid for Σ': With the tree structure on the shearlet index set we can define

$$\Sigma_N'^t := \left\{ \sum_{\mu \in \mathcal{T}} c_\lambda \tilde{\sigma}'_\lambda \colon \mathcal{T} \subset \Lambda \text{ is tree, and } |\mathcal{T}| \leqslant N \right\}$$

and obtain the following result.

Theorem 24. *There exist compactly supported functions φ', $\psi'^{(0)}$, $\psi'^{(1)}$ such that with Σ' as in* (6) *and*

$$t'_N(f) := \inf_{g \in \Sigma_N'^t} \|f - g\|_2.$$

we have for any $\varepsilon > 0$

$$\sup_{f \in \mathcal{F}} t'_N(f) \lesssim N^{-1+\varepsilon}.$$

Remark 25. We would like to remark that Theorem 24 could also be proven more directly by using the results in [27] instead of our results on localization. In particular Remark 13 still holds in the compactly supported case.

4. Applications in image coding

The near-optimality of tree approximation leads to a near-optimal encoding strategy in the same way as in [7] for wavelets. An encoding scheme for $\mathcal{F}$ consists of an *encoder* E which maps an $f \in \mathcal{F}$ to a bitstream $E(f)$, i.e. a sequence of zeros and ones. A *decoder* maps a bitstream onto a function $f \in L_2([0,1]^2)$.

The *distortion* of the encoding/decoding pair (E, D) is defined as

$$d(E, D) := \sup_{f \in \mathcal{F}} \left\| f - D\left(E(f)\right)\right\|_2. \tag{29}$$

For an encoder E we define its *runlength* as

$$M(E) := \sup_{f \in \mathcal{F}} \left|E(f)\right|,$$

where $|E(f)|$ denotes the length of the bitstream $E(f)$. A general encoding/decoding scheme for wavelets is constructed in [7]. The main property that is used is the fact that a general tree can be encoded much less expensively than an unstructured set of indices, *provided that the number of roots in the tree is uniformly bounded*; this is shown in [7, Lemma 6.1]. Therefore, in order to directly apply the results and constructions of [7, Section 6] for constructing good shearlet coding procedures for $\mathcal{F}$, it is essential to establish the fact that the set

$$\mathcal{D}_0 := \left\{\lambda \in \Lambda_{-1} \colon \exists f \in \mathcal{F},\ \lambda' \preccurlyeq \Lambda \colon \left\langle f, \sigma'_{\lambda'} \right\rangle \neq 0\right\}$$

of possible roots is finite. Fortunately, this is the case if the shearlet frame consists of compactly supported functions:

Lemma 26. *If φ', $\psi'^{(0)}$, $\psi'^{(1)}$ are compactly supported and Σ' is constructed as in* (6)*, then* $\operatorname{card} \mathcal{D}_0 < \infty$.

Proof. We show that for all $m \in \mathbb{Z}^2$, there exists a bounded set D in $\mathbb{Z}^2$ such that for all $\lambda \preccurlyeq m$ we have $\operatorname{supp} \sigma'_\lambda \subset m + D$. Since all $f \in \mathcal{F}$ are supported in $[0,1]^2$, this implies that only a finite number of indices $m \in \Lambda_{-1}$ can occur as possible root. For any $\lambda = (j,l,k,d) \in \Lambda$ it is not hard to see that the compact support of the basis functions implies that $\operatorname{supp} \sigma'_\lambda \subset A_d^{-j} B_d^{-l} k + 2^{-j} B$, where B is some bounded set in $\mathbb{R}^2$. We will now write A_λ for the dilation matrix $B_d^l A_d^j$ associated with an index $\lambda = (j,l,k,d)$. The children of m in Λ_0 are given by all indices $\lambda_0 = (0, l_0, k_0, d_0)$ with $k_0 \in B_{d_0}^{l_0} A_{d_0}^0 m$. We shall now drop the subscript d for the matrices A, B and $\mathcal{E}$. The children of m in Λ_1 are given by all indices $\lambda_1 = (1, l_1, k_1, d_1)$ with $k_1 \in B^\nu A k_0 + B^\nu \mathcal{E}$, where $\nu \in \{0,1\}$ and $k_0 \in B^{l_0} A^0 m$ for some l_0 and therefore $k_1 \in A_{\lambda_1} m + A_{\lambda_1} A_{\lambda_0}^{-1} A^{-1} \mathcal{E}$. Iterating this argument shows that $\lambda_n \in \Lambda_n$ is a child of m only if $k_n \in A_{\lambda_n}(m + \sum_{i=2}^{n+1} A_{\mu_i}^{-1} \mathcal{E})$ with some indices $\mu_i \in \Lambda_i$. An elementary computation shows that $\|A_{\mu_i}^{-1}\| \lesssim 2^{-i}$ uniformly for all $\mu_i \in \Lambda_i$. It follows that for $\lambda_n \in \Lambda_n$ we have $\operatorname{supp} \sigma'_{\lambda_n} \subset \bigcup_{e \in \mathcal{E}} m + \sum_{i=2}^{n+1} A_{\mu_i}^{-1} e + 2^{-n} B \subset m + \sum_{i \in \mathbb{N}} 2^{-i} [0,4]^2 + B$. It follows that for all children λ of m we have $\operatorname{supp} \sigma'_\lambda \subset m + D$ with a bounded set D. This proves the assertion. $\square$

Moreover, by Theorem 24, the conclusion of Theorem 6 remains valid for compactly supported shearlet frames.

Using the fact that the set $\mathcal{D}_0$ of roots is finite, we can perform the exact same encoding construction as in [7, Section 6] and construct an encoder E_N which has length $M(E_N) \lesssim 2^{(2/3+\varepsilon)N}$ for all $\varepsilon > 0$ and $N \in \mathbb{N}$ and a decoder D_N with

$$d(E_N, D_N) \lesssim 2^{-(2/3-\varepsilon)N}.$$

It follows that

$$d(E_N, D_N) \lesssim M(E_N)^{-1+\varepsilon}$$

for all $\varepsilon > 0$, a result that is optimal if we disregard the arbitrarily small ε, compare [16].

Remark 27. For the convenience of the reader we sketch a simpler encoding–decoding procedure which is sufficient for our purposes: Let $N \in \mathbb{N}$ and $f \in \mathcal{F}$ be given. We describe the encoding procedure:

1. Consider the coefficient set $\mathcal{T}(f, N^{-2/3})$. By (13) we have $|\mathcal{T}(f, N^{-2/3})| \lesssim N^{1+\varepsilon}$ for $\varepsilon > 0$ arbitrary.
2. Since $\mathcal{T}(f, N^{-2/3})$ forms a subtree of the index set Λ which by Lemma 26 has only finitely many roots, we need $\lesssim N^{1+\varepsilon}$ bits to store the kept indices (this follows from [7, Lemma 6.1]).
3. Now we round the kept frame coefficients up to precision $N^{-2/3}$. Thus, for each coefficient we need order $\log(N)$ bits which sums up to order $N^{1+\varepsilon}$ bits for $\varepsilon > 0$ arbitrary.
4. Store the index set $\mathcal{T}(f, N^{-2/3})$ together with the rounded frame coefficients. This requires order $N^{1+\varepsilon}$ bits in total, $\varepsilon > 0$ arbitrary.

The decoding works in the obvious way, namely by reconstructing from the kept indices and the quantized frame coefficients using the dual frame $\tilde{\Sigma}'$. Using (15) one can show that this reconstruction gives an error of order $N^{-1+\varepsilon}$, which is what we want.

Having a close-to-optimal bit rate coding procedure allows us to draw some conclusions regarding the Kolmogorov entropy of $\mathcal{F}$. We equip $\mathcal{F}$ with the metric inherited from $L_2(\mathbb{R}^2)$. It is not difficult to see that $\mathcal{F}$ is contained in a compact subset of $L_2(\mathbb{R}^2)$. For any $\nu > 0$ there exists a minimal number N_ν such that $\mathcal{F}$ can be covered by N_ν balls with diameter ν. The Kolmogorov ν-entropy H_ν is defined by

$$H_\nu := \log N_\nu.$$

Corollary 28. *For any $\varepsilon > 0$ the Kolmogorov ν-entropy satisfies*

$$H_\nu \lesssim \nu^{-1-\varepsilon}.$$

Proof. Using the encoding/decoding pair described above, we can consider the image of $\mathcal{F}$ under the mapping E_N which has cardinality $\lesssim 2^{M(E_N)}$. Now consider the system of balls with midpoints $\{D_N(E_N(f)):\ f \in \mathcal{F}\}$ and radius $\sim M(E_N)^{-1+\varepsilon}$. By the fact that $d(E_N, D_N) \lesssim M(E_N)^{-1+\varepsilon}$, it follows that this system is a covering of $\mathcal{F}$. On the other hand, the number of elements in this covering is $2^{M(E_N)}$ and therefore $H_{M(E_N)^{-1+\varepsilon}} \lesssim M(E_N)$. This proves the statement. $\square$

Of course there exist several other methods to bound the Kolmogorov entropy of $\mathcal{F}$, see e.g. [17,30]. However, the method outlined in this section provides a particularly simple proof. Also the coding procedure which we presented is very simple: It is based on simple hard thresholding of the frame coefficients of f with respect to a *nonadaptive* frame. This stands in contrast to other adaptive methods like for instance bandlets [30]. We also want to emphasize the important point that encoding and decoding based on shearlets can be done in log-linear time, a property not shared by adaptive methods.

Acknowledgments

The author would like to thank Christoph Schwab for useful discussions, as well as the anonymous referees for raising important points which greatly improved the presentation. A part of the results of the present work has been announced in [20].

References

[1] T. Berger, Rate Distortion Theory, Prentice Hall, 1970.
[2] E. Candes, L. Demanet, The curvelet representation of wave propagators is optimally sparse, Comm. Pure Appl. Math. 58 (2004) 1472–1528.
[3] E. Candes, D. Donoho, Curvelets – a surprisingly effective nonadaptive representation for objects with edges, in: Curves and Surfaces, Vanderbilt University Press, 1999.
[4] E. Candes, D. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities, Comm. Pure Appl. Math. 57 (2002) 219–266.
[5] E. Candes, D. Donoho, Continuous curvelet transform: II. Discretization and frames, Appl. Comput. Harmon. Anal. 19 (2) (2005) 198–222.
[6] O. Christensen, An Introduction to Frames and Riesz Bases, Birkhäuser, 2003.
[7] A. Cohen, W. Dahmen, I. Daubechies, R. Devore, Tree approximation and optimal encoding, Appl. Comput. Harmon. Anal. 11 (2001) 192–226.
[8] A. Cohen, R. DeVore, P. Petrushev, H. Xu, Nonlinear approximation and the space $BV(\mathbb{R}^2)$, Amer. J. Math. 121 (1999) 587–628.
[9] A. Cohen, N. Dyn, F. Hecht, J.-M. Mirebeau, Adaptive multiresolution analysis based on anisotropic triangulations, Technical report, 2009.
[10] R. Coifman, M. Wickerhauser, Entropy based algorithms for best basis selection, IEEE Trans. Inform. Theory 32 (March 1992) 712–718.
[11] L. Demaret, N. Dyn, A. Iske, Image compression by linear splines over adaptive triangulations, Signal Process. 86 (2006) 1604–1616.
[12] R. DeVore, Nonlinear approximation, Acta Numer. (1998) 51–150.
[13] R. DeVore, G. Lorentz, Constructive Approximation, Springer, 1993.
[14] M. Do, M. Vetterli, The contourlet transform: An efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (2005) 2091–2106.
[15] D. Donoho, Unconditional bases and bit-level compression, Appl. Comput. Harmon. Anal. 3 (1996).
[16] D. Donoho, Wedgelets: Nearly minimax estimation of edges, Ann. Statist. 27 (1998) 353–382.
[17] D. Donoho, Sparse components of images and optimal atomic decompositions, Constr. Approx. 17 (1999) 353–382.
[18] D. Donoho, G. Kutyniok, Sparsity equivalence of anisotropic decompositions, Technical report, 2010.
[19] G. Easley, W.-Q. Lim, D. Labate, Sparse directional image representations using the discrete shearlet transform, Appl. Comput. Harmon. Anal. 25 (2008) 25–46.
[20] P. Grohs, Tree approximation and optimal image coding with shearlets, in: Proceedings of the SAMPTA 2011 Conference (Singapore), in press.
[21] K. Guo, D. Labate, Optimally sparse multidimensional representation using shearlets, SIAM J. Math. Anal. 39 (1) (2008) 298–318.
[22] K. Guo, D. Labate, Representation of Fourier integral operators using shearlets, J. Fourier Anal. Appl. 14 (3) (2008) 327–371.
[23] K. Guo, D. Labate, W.-Q. Lim, G. Weiss, E. Wilson, Wavelets with composite dilations and their MRA properties, Appl. Comput. Harmon. Anal. 20 (2) (2006) 202–236.
[24] P. Kittipoom, G. Kutyniok, W.-Q. Lim, Construction of compactly supported shearlet frames, Constr. Approx. (2011), in press.
[25] G. Kutyniok, J. Lemvig, W.-Q. Lim, Compactly supported shearlets, in: Approximatin Theory XIII, San Antonio, TX, 2010, Springer, 2011.
[26] G. Kutyniok, J. Lemvig, W.-Q. Lim, Shearlets and optimally sparse approximations, in: Shearlets: Multiscale Analysis for Multivariate Data, Springer, 2011.
[27] G. Kutyniok, W.-Q. Lim, Compactly supported shearlets are optimally sparse, J. Approx. Theory 163 (2011) 1564–1589.
[28] G. Kutyniok, W.-Q. Lim, Shearlets on bounded domains, in: Approximation Theory XIII, San Antonio, TX, 2010, Springer, 2011.
[29] D. Labate, W.-Q. Lim, G. Kutyniok, G. Weiss, Sparse multidimensional representation using shearlets, in: Proceedings of the SPIE, 2005, pp. 254–262.
[30] E. LePennec, S. Mallat, Sparse geometric image representation with bandlets, IEEE Trans. Image Process. 14 (2005) 423–438.
[31] E. Ott, Chaos in Dynamical Systems, Cambridge University Press, 1993.
[32] H. Smith, A Hardy space for Fourier integral operators, J. Geom. Anal. 8 (1998) 629–653.